REMARKS

In the Final Office Action, Claims 1, 3-6, 8-10, 12-15, 17 and 18 are pending and stand rejected. In response, Claims 1 and 10 are amended, Claims 6 and 15 are canceled and no claims are added. The Applicants respectfully request reconsideration in view of the following remarks and amendments.

I.   Claims Rejected Under 35 U.S.C. § 103

Claims 1, 3-6, 8-10, 12-15, and 17-18 stand rejected under 35 U.S.C. § 103(a) as being unpatentable over U.S. Patent No. 7,024,400 issued to Tokuda et al. ("Tokuda") in view of U.S. Patent Publication No. 2003/0221163 to Glover ("Glover") and U.S. Patent Publication No. 2001/0037324 to Agrawal ("Agrawal"). Applicants respectfully traverse this rejection.

Claim 1 recites:

1.   A sentence classification device characterized by comprising:
    ... wherein the classification generation module comprises a virtual representative document generation module for generating a virtual representative document, for each cluster on a transformed DT matrix, from a term of each document belonging to the cluster;
        large classification generation module for generating a large classification of documents from each document in a bottom-up manner by repeatedly performing hierarchical clustering processing of setting a DT matrix generated by said DT matrix generation module in an initial state, causing said virtual representative document generation module to generate a virtual representative document for each cluster on a transformed DT matrix generated from the DT matrix by said DT matrix transformation module, generating a new DT matrix used for next hierarchical clustering processing by adding the virtual representative document to the transformed DT matrix and deleting documents belonging to the cluster of the virtual representative document from the transformed DT matrix, and outputting, for said each cluster, information associated with the documents constituting the cluster as large classification data;
        term list edition module for adding or deleting an arbitrary term with respect to the term list; and
        index generation module for making said DT matrix generation module generate DT matrices by using term lists before and after edition by said term list edition module, and generating and outputting an index indicating validity of the edition from the DT matrices.

The present invention is directed to unsupervised classification that clusters unclassified common texts (not limited to Web documents) into a hierarchical structure (taxonomy) in an unsupervised manner. A principal concept of the present invention is to perform clustering processing through repeated execution of the graph theory called DM decomposition. In the present invention, a process of transforming (degenerating) a graph matrix in this repeated processing is referred to as "generating a virtual document (virtual representative document)."

In contrast, Tokuda is directed to a supervised document classification procedure using a differential latent semantics index (DSLI) approach in which a classification label is provided as a supervisor. Tokuda teaches to first construct a model for a collection of labeled documents having labels indicating their classifications by using a differential latent semantics index (DLS() approach in which a classification label is provided as a supervisor. Tokuda teaches to first construct a model for a collection of labeled documents having labels indicating their classifications by using a DLSI method derived from a latent semantics index (LSI) space-based approach. Tokuda also teaches to use the thus built model to classify a collection of unlabeled unclassified documents.

Moreover, regarding Agrawal, the reference Agrawal is directed to a supervised classification which determines where in a hierarchical structure an unclassified text is to be allocated using as a supervisor those texts that are previously arranged in the hierarchical structure. The principal concept of Agrawal is based on a statistical model that calculates a Fisher value (well-known in statistics) from words or information contained in the text to then determine, depending on the magnitude of the Fisher value, where the text is to be allocated in a hierarchical structure provided as a supervisor.

Regarding Glover, Glover is directed to a supervised classification that determines into which category an unclassified Web page is to be classified using a predetermined category as a supervisor while using information of sentences containing hyperlink letters (anchortexts) of the unclassified Web page document. In Glover, sentences containing hyperlinks to classification target Web documents (e.g., Yahoo) from a Web document (referring to Yahoo) hyperlinking to the classification target Web document (e.g., Yahoo) are first extracted and then put together to generate virtual documents. By subjecting such virtual documents to classification processing

using an ordinary classifier (e.g., SVM (Support Vector Machine)), an improvement in classification accuracy can be expected over cases without use of virtual documents (e.g., with classification target Web documents only).

Claim 1 features an unsupervised classification technique, distinctly different from Tokuda and the other references in principal concept of classification. Also, the supervised classification of Agrawal requires initial generation of supervisor information by training documents and then classifying unclassified texts based on the thus generated supervisor information, thus involving a two-step process. Further, Agrawal employs a classification method using a statistical model based on Fisher values, a method distinctly different from the method of the present invention based on repetitive execution of DM decomposition.

The supervised classification of Glover initially generates a model (document vector) from a positive instance (positive document collection) belonging to a category and then classifies unclassified Web documents using the thus generated model. Glover generates virtual documents from hyperlinks of Web documents. Therefore, no virtual documents will be generated without hyperlinks. The present invention does not employ hyperlinks.

In addition, the Examiner has conceded that Tokuda does not explicitly teach the details of "classification generation module," added to the original claim 1 when responding to the preceding Office Action, namely "a virtual representative document generation module" and "a large classification generation module" (original claim 7). In this regard, although the Examiner asserts that the "virtual representative document generation module" is taught in Glover, Glover neither teaches nor suggests generating a virtual representative document for each cluster on a transformed DT matrix. By the same token, the Examiner asserts that the "large classification generation module" is taught in Agrawal, but Agrawal neither teaches nor suggests repeating the large classification outputting in association with new DT matrix generation for each cluster.

Further, the Examiner has admitted that Tokuda does not explicitly teach the "term list edition module" (original claim 6) that is herein added to claim 1. Also in this connection, the Examiner has referred to Agrawal. Furthermore, as to the "index generation module," the idea of generating an index indicating a usefulness in association with the DT matrix generation is not taught in any of the cited three references.

Finally, Tokuda does not teach the component elements featuring the present invention at least in association with the DT matrix. These component elements are not taught by either Glover or Agrawal.

For each of the above reasons, Claim 1, and all claims which depend from Claim 1, are patentable over the cited references. Therefore, please reconsider and withdraw the § 103(a) rejection of Claims 1, 2-6, and 8-9.

Each of the Applicants other independent claims include limitations similar to those discussed above. Therefore, all of the Applicants other independent claims, and all claims which depend on them, are patentable over the cited art for similar reasons. Consequently, Applicants respectfully request that the Examiner reconsider and withdraw the §103 rejection of Claims 10, 12-15 and 17-18.

DEPENDENT CLAIMS

In view of the above remarks, a specific discussion of the dependent claims is considered to be unnecessary. Therefore, Applicant's silence regarding any dependent claim is not to be interpreted as agreement with, or acquiescence to, the rejection of such claim or as waiving any argument regarding that claim.
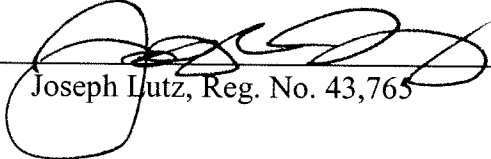
CONCLUSION

In view of the foregoing, it is submitted that the pending claims patentably define the subject invention over the cited references of record, and are in condition for allowance and such action is earnestly solicited at the earliest possible date. If the Examiner believes a telephone conference would be useful in moving the case forward, he is encouraged to contact the undersigned at (310) 207-3800.

If necessary, the Commissioner is hereby authorized in this, concurrent and future replies, to charge payment or credit any overpayment to Deposit Account No. 02-2666 for any additional fees required under 37 C.F.R. §§1.16 or 1.17, particularly, extension of time fees.

Respectfully submitted,
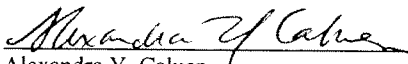
BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN LLP

Dated: May 5, 2008

Joseph Lutz, Reg. No. 43,765

1279 Oakmead Parkway
Sunnyvale, California 94085-4040
Telephone (310) 207-3800
Facsimile (408) 720-8383